MULTILEVEL MODELLING NEWSLETTER

Centre for Multilevel Modelling
Mathematical Sciences
Institute of Education, University of London
20 Bedford Way, London WC1H 0AL, ENGLAND
Web site http://multilevel.ioe.ac.uk
Enquiries about newsletter to Ian Plewis
E-mail: i.plewis@ioe.ac.uk
Tel: +44 (0) 20 7612 6688
Fax: +44 (0) 20 7612 6572

Vol. 13 No. 2 **December, 2001**

Forthcoming Workshops

8-10 April 2002, a three-day introductory workshop to multilevel modelling using *MLwiN* will take place in the University of Bristol.

Enquiries to Jean Flowers at Graduate School of Education, 35 Berkeley Square, Bristol BS8 1HJ, United Kingdom. Tel: +44 (0) 117 928 7059, Fax: +44 (0) 117 925 5412, email: jean.flowers@bristol.ac.uk.

8-10 May 2002, a three-day introductory workshop to multilevel modelling using *MLwiN* will take place in the Institute of Education, University of London.

This workshop can be booked on-line: http://multilevel.ioe.ac.uk/support/workshop.html.

Enquiries to Amy Burch at Mathematical Sciences, Institute of Education, 20 Bedford Way, London WC1H OAL, United Kingdom. Tel: +44 (0) 20 7612 6688, Fax: +44 (0) 20 7612 6572, email: a.burch@ioe.ac.uk.

If you plan to run any workshops using *MLwiN*, please notify Amy Burch and she will advertise these workshops on the multilevel web site.

Also in this issue
News about conferences
The effect on variance component estimates of ignoring a level in a multilevel model
A non-parametric random coefficient approach: the latent class regression model
A review of the Mplus statistical package
A review of 'Multilevel Modelling of Health Statistics'
Some new references on multilevel modelling

53rd Session of the International Statistical Institute (ISI), Seoul, 22-29 August 2001

This conference included an invited session on Multilevel Models for Survey Design and Analysis. There were three papers:

1. Sample Size Considerations for Multilevel Surveys (Michael P. Cohen; Bureau of Transportation Statistics, USA)

2. Selection Effects in an Analysis of Contraceptive Discontinuation in Morocco: An Application of a Multiprocess Multilevel Model. (Fiona Steele; Institute of Education)

3. Multilevel Modelling under Informative Probability Sampling. (Danny Pfeffermann; Hebrew University, Israel: Fernando Moura; Federal University of Rio de Janeiro, Brazil: and Pedro Nascimento Silva; IBGE, Brazil).

In his paper, Cohen discussed how analysts intending to use multilevel modelling need to take this into account at the survey design stage. He described how traditional methods for sample size determination in two-stage surveys may be adapted to determine the most efficient sample allocation (i.e. the optimal numbers of level two units, and level one units within level two units) subject to fixed cost constraints. Typically, sample size determination involves minimising the variance of the parameter estimate of interest, e.g. a regression coefficient or intra-class correlation coefficient. Cohen discussed ways in which these methods might be extended to multi-purpose surveys where there are many parameters of interest.

Steele considered multilevel models for situations where explanatory variables might be correlated with random effects at one or more levels. In the application described, the outcome variable of interest is the duration to discontinuation of contraceptive use and a key explanatory variable is the source of contraceptive supply (a private or public facility). In this case, there may be unobserved individual- and community-level factors that influence both a couple's choice of service provider and their probability of discontinuation, leading to correlation between individual- and community-level random effects and source of supply. Steele discussed how multilevel multiprocess models can be used to model jointly the processes of contraceptive discontinuation and choice of provider.

Pfeffermann and his co-authors considered multilevel models for survey data where clusters and/or final sample units have been selected with unequal selection probabilities. If selection probabilities are related to the outcome variable then the sampling process is said to be informative. For example, if clusters are selected with probability proportional to size and the outcome variable is inversely related to population density, the sample will tend to contain large clusters with low values on the outcome variable and will therefore not be representative of the population. A model-based approach for multilevel analysis which accounts for informative sampling was described. The authors presented results from a simulation experiment designed to assess the performance of their method and to demonstrate the impact of ignoring the sampling design.

Short versions of the papers appear in the Bulletin of the ISI and may be downloaded from: http://www.nso.go.kr/isi2001/.

Longer versions of invited papers from this session and other sessions organised by the International Association of Survey Statisticians will appear in a volume to be published by the Australian Bureau of Statistics.

Fiona Steele
Institute of Education
University of London

9th Annual Meeting of the Belgian Statistical Society, Oostende, 12 and 13 October 2001

At the 2001 meeting of the Belgian Statistical Society, organised by the University Centre for Statistics of the Katholieke Universiteit Leuven, there was a notable presence of invited talks and contributed papers on mixed/multilevel models and applications. Ten of the 39 presentations in the plenary and parallel sessions had their main focus on these kinds of models. This is remarkable since the other contributions ranged from Asymptotic Theory, Bayesian Statistics, Robust Statistics to Machine Learning, and the Belgian Statistical Society has strong research groups in each of these areas.

Douglas Bates, who co-authored a book on mixed-effects models in S and S-Plus together with José Pinheiro, was the first invited speaker. In his tutorial on mixed-effects models in practice Professor Bates set the stage for the remaining presentations in this domain. He discussed linear and non-linear mixed-effects models and how they are implemented in their NLME package for the S or R-language, although he also mentioned the NLMIXED procedure of SAS, and the HLM and *MLwiN* software packages. Other interesting presentations on the first day were:

- Mixtures of non-linear mixed models for the classification of longitudinal profiles (by Steffen Fieuws)
- Using mixed models to detect differential item functioning across multiple groups (by Jerry Welkenhuysen-Gybels)
- Types of design matrices for test data with a componential design (by Paul De Boeck)
- Non-linear mixed models and clinical trial simulation (by Filip De Ridder)

On the second day, Min Yang was our invited speaker. She reviewed the research of the Multilevel Models Project on meta-analysis using multilevel models. This use of multilevel models was illustrated in combining aggregated binary data, individual binary data, aggregated continuous data, and mixtures of aggregated and individual level continuous data. Other presentations:

- Triple-goal estimates in linear mixed models (by Wendim Ghidey)
- The linear logistic test model with random effects (by Frank Rijmen)
- Parametric and nonparametric bootstrap methods for meta-analysis (by Wim Van den Noortgate)
- Modelling interviewer effects in panel surveys: an application (by Jan Pickery)

also had an almost exclusive mixed/multilevel focus.

There seems to be a strong (and even still growing?) interest in mixed and multilevel models in Belgian biomedical, social, behavioural and educational research institutes and universities. For more information on this meeting, the selected contributions, and addresses for correspondence please refer to:
http://www.kuleuven.ac.be/ucs
or
http://www.kuleuven.ac.be/ucs/BSS2001

Patrick Onghena
Department of Educational Sciences
Katholieke Universiteit Leuven

The effect on variance component estimates of ignoring a level in a multilevel model

Dougal Hutchison[1], Michael Healy[2]
[1]National Foundation for Educational Research, Slough, UK
[2]Institute of Education, University of London
d.hutchison@nfer.ac.uk

Table 1 shows the results of variance components analyses of some unpublished data on attainment scores in mathematics in English schools.

Table 1
Variance components for mathematics attainment scores.

	Three level	Two level
Between schools	3.7	5.0
Between classes	8.6	-----
Between pupils	7.0	13.3

Total number of pupils: 2718

Mean number of classes per school: 5.09; range 2-10.

The first column in the table shows the estimates from a three level model, pupils within classes within schools; the second column is from a two level model in which the grouping of pupils into classes has been ignored. As would be expected, the between pupils component has increased since this now has to include the between classes differences. However, it will be seen that the between schools component has also increased. This might be regarded as surprising since there has been no change in the mean performance of each school. Clarifying the issue requires some algebra.

Suppose for simplicity that we have a balanced situation with s schools, q classes in each school and p pupils in each class. The three level analysis of variance is as follows:

	df	SS	MS	Expected MS
Schools	s-1	A	a	$pq\alpha+p\beta+\gamma$
Classes	$s(q$-1)	B	b	$p\beta+\gamma$
Pupils	$sq(p$-1)	C	c	γ
Total	sqp-1	D		

where α, β and γ are the variance components. The between schools variance is estimated by $(a - b)/pq$.

If classes are ignored, the two level ANOVA is:

	df	SS	MS	Expected MS
Schools	s-1	A	a	$pq\alpha'+\gamma'$
Pupils	$s(pq$-1)	E	e	γ'
Total	sqp-1	D		

A and D are unchanged, so E = B + C, $e = (B + C)/\{s(pq-1)\}$ and the between schools variance component is estimated by $\alpha' = (a - e)/pq$. Then
$\gamma' = \gamma + \{(pq - p)/(pq - 1)\}\beta$
and
$\alpha' = \alpha + \{(p - 1)/(pq - 1)\}\beta$.
Thus the between schools component is also increased by an amount depending upon the ignored between-class variability. The same increase seems likely to occur in unbalanced situations although it is difficult to generalise about the size of the effect.

The phenomenon seems to be a reflection of the well-known fact that the estimated error variance of higher level means tend to be too small when the presence of a hierarchy at lower levels is ignored. In the three level model, the error variance of a school mean is

$$\frac{\beta}{q} + \frac{\gamma}{pq} = \frac{1}{pq}(p\beta + \gamma)$$

In the two level model this becomes

$$\frac{\gamma'}{pq} = \frac{1}{pq}(p.\frac{q-1}{pq-1}\beta + \gamma)$$

where the coefficient of β has been multiplied by a factor which is less than 1. Just as the precision of the estimated school means is exaggerated by the incorrect model, so the estimate of their true variability is similarly exaggerated. Since the observed school level variance is a combination of the true variance and the error variance, this underestimate of the error variance leads to a corresponding overestimate of the true variance.

A non-parametric random coefficient approach: the latent class regression model

Jeroen K. Vermunt and Liesbet A. van Dijk
Department of Methodology and Statistics, Tilburg University, Netherlands
l.a.vandijk@kub.nl

Introduction

Most of the work in the area of random coefficient modelling has focused on parametric methods in which the random coefficients are assumed to come from a known distribution, typically a multivariate normal distribution (Bryk and Raudenbush, 1992; Goldstein, 1995; Hedeker and Gibbons, 1996; Agresti et al., 2000). This paper presents latent class (LC) analysis as a non-parametric random coefficient model. Advantages of our proposed LC regression model are that less restrictive assumptions are made about the distribution of the random effects and that any model of the generalised linear modelling (GLM) family can be dealt with without increasing computation time. User friendly software with an SPSS-like interface is available to apply the proposed method (Vermunt and Magidson, 2000; www.LatentGold.com).

In the next section, we describe the LC regression model and compare it with the parametric random coefficient model. We then discuss parameter estimation by maximum likelihood (ML) followed by an application using an empirical dataset. We end with some final remarks.

The latent class regression model

Let i denote a level-1 case within the level-2 case j. Let x denote a level-1 predictor and w a level-2 predictor. The general parametric two level model can be defined as follows:

$$\begin{aligned} \eta_{ij} &= \sum_{q=0}^{Q} \beta_{qj} x_{qij} + e_{ij}, \\ \beta_{qj} &= \sum_{s=0}^{S} \gamma_{qs} w_{sj} + u_{qj}, \end{aligned} \qquad (1)$$

where $\mathbf{u}_j \sim N(\mathbf{0},\mathbf{T})$. The distribution of e_{ij} can be any function belonging to the exponential family. Note that x_{0ij} and w_{0j} equal 1, which makes β_{0j} and γ_{q0} intercepts.

Using the same notation as in Equation (1) and indexing the latent classes by k, the LC regression model can be defined as follows:

$$\begin{aligned} \eta_{ij} &= \sum_{q=0}^{Q} \beta_{qk} x_{qij} + e_{ij}, \\ \beta_{qk} &= \sum_{s=0}^{S} \gamma_{qs} w_{sj} + u_{qk}, \end{aligned} \qquad (2)$$

where the distribution of $\mathbf{u}_k$ is unspecified, that is, $p(\mathbf{u}_k) = \pi_k$. For identification and comparability with the parametric two level model, we set $\sum_{k=1}^{K} u_{qk}\pi_k = 0$. Note that in the standard formulation of the LC regression model, the first equation in (2) suffices.

Comparison of the LC model described in (2) with the parametric two level model of (1) shows that rather than having a separate set of regression coefficients for each individual coming from a multivariate normal distribution, we assume that there exists a finite number of subgroups with different regression coefficients (Wedel and DeSarbo, 1994). This can be seen as a fundamental difference between the two models, especially if one is interested in identifying latent classes. However, the LC regression model can also be seen as a non-parametric two level model; that is, as a two level model in which no assumptions are made about the distributional form of the random effects. With the maximum number of identifiable latent classes, the distribution may be interpreted as a non-parametric distribution (Laird, 1978; Rabe-Hesketh et al., 2001). In practice, however, we will stop increasing the number of latent classes when the model fit no longer improves. It should be noted that the current LC regression model cannot deal with more than two levels.

The conceptual equivalence between the LC regression and the two level model becomes even clearer if we compute the second-order moments of the random coefficients from the standard latent class parameters. In a model without cross-level interactions, these are obtained by:

$$\tau_{qq'} = \sum_{k=1}^{K} u_{qk}u_{q'k}\pi_k = \sum_{k=1}^{K} \left(\beta_{qk} - \gamma_{q0}\right)\left(\beta_{q'k} - \gamma_{q'0}\right)\pi_k, \tag{3}$$

where $\gamma_{q0} = \sum_{k=1}^{K} \beta_{qk}\pi_k$.

Equation (3) shows that the results of a LC regression analysis can be summarised in the same way as in a two level model; that is, in terms of a fixed and random part.

Parameter estimation

LC regression models are usually estimated by maximum likelihood (ML). The likelihood contribution of level-two unit j equals

$$f(\mathbf{y}_j|\mathbf{x}_j, \mathbf{w}_j) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_j|\mathbf{x}_j, \mathbf{w}_j) = \sum_{k=1}^{K} \pi_k \prod_i f_k(y_{ij}|\mathbf{x}_j, \mathbf{w}_j),$$

where K is the number of latent classes and $f_k(y_{ij}|\mathbf{x}_j, \mathbf{w}_j)$ is a class-specific density. This density can be any function belonging to the exponential family.

The most popular algorithm to solve the ML estimation problem is the EM algorithm. The Latent GOLD software (Vermunt and Magidson, 2000) that was used for the example reported in the next section combines EM with Newton-Raphson. More precisely, the estimation process starts with a number of EM iterations and switches to Newton-Raphson when the relative change in the parameters is small. Local optima are avoided by using multiple sets of random starting values. Other software packages that can be used to estimate LC regression models are LEM (Vermunt, 1997), GLIMMIX (Wedel and DeSarbo, 1994), and GLLAMM (Rabe-Hesketh et al., 2001).

Contrary to the non-parametric method, parameter estimation in parametric random coefficient models can become quite complex and time consuming when the distribution of the dependent variable is non-normal, such as with discrete response variables. Approximation methods to deal with the complicated integrals in the likelihood equations are numerical integration, Monte Carlo integration, and first- or second-order Taylor expansion of the link function (Agresti et al., 2000). It should be noted that the quite popular quadrature approximation of the likelihood that is used in the MIXOR (Hedeker and Gibbons, 1996) and GLLAMM (Rabe-Hesketh et al, 2001) packages is equivalent to using a LC model with many latent classes, where the location and weights of the classes are fixed rather than estimated from the data.

An application

In order to compare the results of parametric and non-parametric random coefficient models, we used a dataset obtained from the data library of the Centre for Multilevel Modelling multilevel.ioe.ac.uk/intro/datasets.html. The data consist of 264 participants in 1983 to 1986 yearly waves from the British Social Attitudes Survey (McGrath and Waterton, 1986). It is a three level dataset: individuals are nested within electoral constituencies and time-points are nested within individuals. We will only make use of the latter nesting, which means that we are dealing with a standard repeated measures model. As was shown by Goldstein (1995), the highest level variance – between constituencies – is so small that it can reasonably be ignored.

The dependent variable is the number of yes responses on seven yes/no questions as to whether it is a woman's right to have an abortion under a specific circumstance. Because this variable is a count with a fixed total, it is most natural to work with a logit link and binomial error function. Individual level predictors in the dataset are religion, political preference, gender, age, and self-assessed social class. In accordance with the results of Goldstein (1995), we found no significant effects of gender, age, self-assessed social class, and political preference. Therefore, we did not use these predictors in the further analysis. The predictors that were used are the level-1 predictor year of measurement (1=1983; 2=1984; 3=1985; 4=1986) and the level-2 predictor religion (1=Roman Catholic, 2=Protestant; 3=Other; 4=No religion).

The non-parametric models were estimated by means of version 2.0 of the Latent GOLD program. Using the elementary statistics computations described in (3), we obtained the multilevel type γ and τ parameters from the standard LC regression output. The parametric models were estimated with quadrature approximation of the likelihood. We used 10 nodes for the random intercept and 6 nodes for random slopes, which with 3 random slopes amounts to having a restricted "latent class" model with 2160 latent classes. The quadrature method was implemented in an experimental version of Latent GOLD. It is, however, not available in version 2.0 of the program.

First, three models without random effects were estimated: an intercept only model (Ia), a model with a linear effect of year (Ib), and a model with year dummies (Ic). Models Ib and Ic also contained the nominal level-2 predictor religion. The test results reported in the first part of Table 1 show that year and religion have significant effects on the dependent variable and that it is better to treat year as non-linear.

Table 1. Test results for the estimated models with the attitudes towards abortion data

Model	Log-Lik.	BIC	Npar
I. No random effects			
a. empty model	-2308.6	4622.8	1
b. time linear	-2215.2	4458.4	5
c. time dummies	-2188.4	4415.8	7
II. Random intercept (time dummies)			
a. parametric (10 nodes)	-1711.8	3468.1	8
b. 2-class	-1754.7	3559.5	9
c. 3-class	-1697.4	3456.2	11
d. 4-class	-1689.5	3451.4	13
e. 5-class	-1689.5	3462.6	15
III. Random intercept and slope (time dummies)			
a. parametric (10, 6, 6, and 6 nodes)	-1695.7	3486.1	17
b. 2-class	-1745.4	3557.8	12
c. 3-class	-1682.7	3460.2	17
d. 4-class	-1656.7	3436.1	22
e. 5-class	-1645.2	3441.0	27

We proceeded by adding a random intercept to Model Ic using the parametric and non-parametric approach described in this paper (Models IIa-IIe). The test results show that both the parametric and the non-parametric random effects models fit better than Model Ic. When using a latent class approach, the model with four classes is the best one in terms of the value of the Bayes Information Criterion (BIC). It can also be seen that the four class model fits much better than the parametric model.

Subsequently, we included random slopes (Models IIIa-IIIe). Within the parametric approach, random slopes did not improve the fit in terms of BIC. In contrast, the LC models with random slopes are better than the models without random slopes. Again the four class model is the best one in terms of BIC. It turns out that this dataset, the more flexible non-parametric approach is better able to capture the individual variation in the slopes than the more restricted parametric method, even with the same number of parameters as in the case of the three class model.

Table 2. Estimates of multilevel parameters for Models Ic, IIa, IId, IIIa, and IIId of Table 1.

Effect	Model Ic	Model IIa[1]	Model IId[2]	Model IIIa[1]	Model IIId[2]
Fixed part					
γ_0	1.50 (0.07)	1.97 (0.13)	*1.89*	2.23 (0.16)	*1.83*
Time					
γ_1 (1983)	-0.13 (0.08)	-0.16 (0.08)	-0.16 (0.08)	-0.35 (0.12)	*-0.13*
γ_2 (1984)	-0.55 (0.07)	-0.68 (0.08)	-0.67 (0.08)	-0.91 (0.11)	*-0.70*
γ_3 (1985)	-0.22 (0.08)	-0.27 (0.08)	-0.26 (0.08)	-0.34 (0.12)	*-0.15*
Religion					
γ_4 (Catholic)	-1.08 (0.10)	-1.07 (0.21)	-1.64 (0.25)	-1.24 (0.31)	-0.95 (0.17)
γ_5 (Protestant)	-0.38 (0.06)	-0.49 (0.19)	-0.22 (0.14)	-0.57 (0.17)	-0.23 (0.11)
γ_6 (Other)	-0.82 (0.08)	-1.12 (0.17)	-0.66 (0.17)	-1.24 (0.20)	-0.52 (0.18)
Random part					
τ_{00}		1.45	*2.05*	2.36	*2.03*
τ_{11}				0.25	*0.19*
τ_{22}				0.42	*0.26*
τ_{33}				0.31	*0.60*
τ_{01}				-0.59	*0.00*
τ_{02}				-0.82	*-0.28*
τ_{03}				-0.29	*-0.10*
τ_{12}				0.30	*0.20*
τ_{13}				0.01	*-0.11*
τ_{23}				0.14	*-0.03*

1. In the quadrature procedure one estimates the Choleski decomposition of T rather than T itself. Our procedure does not therefore yield standard errors for the τ parameters. Standard errors could, however, be obtained by the delta method.
2. We do not report standard errors for the (italicised) parameters, which are derived from the Latent GOLD output using Equation (3). These standard errors could, however, be obtained by the delta method. It should be noted that Latent GOLD provides standard errors, as well as two types of Wald tests for the standard LC regression parameters.

Table 2 reports the multilevel parameter estimates for Models Ic, IIa, IId, IIIa, and IIId. As far as the fixed part is concerned, the substantive conclusions are similar in all five models. The attitudes are most positive at the last time point (reference category) and most negative at the second time point. Furthermore, the effects of religion show that people without religion (reference category) are most in favour and Roman Catholics and Others are most against abortion. Protestants have a position that is close to the no-religion group. A difference between the parametric and non-parametric models is that in the former, Others are as extreme as Roman Catholics, while in the latter it is clearly an intermediate group.

Also the random parts of the parametric models are quite similar. Some differences are that the variance of the intercept is higher in Model IId than in Model IIa. The intercept, time-point one and time-point two variances are somewhat higher in Model IIIa than in Model IIId, but the time-point three variance is much lower. Furthermore, the covariances are much higher in the parametric than in the non-parametric model.

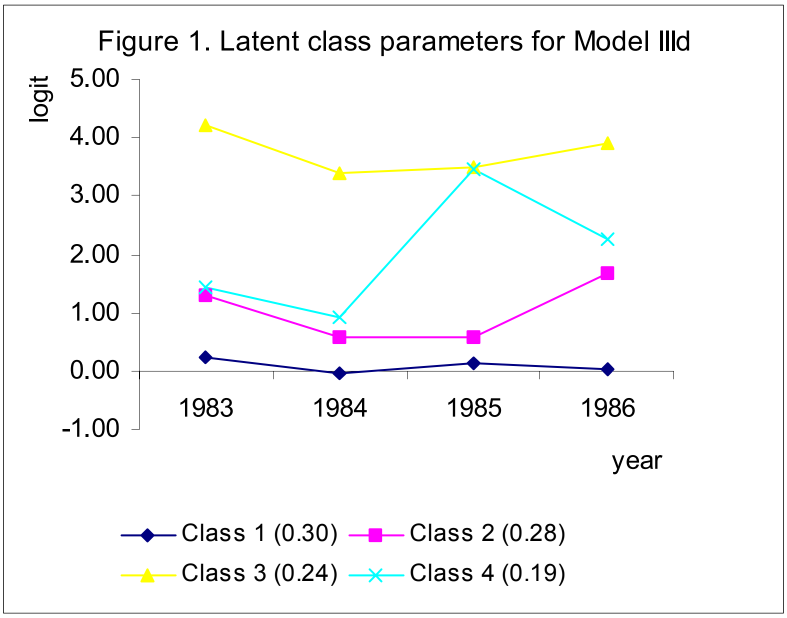

Figure 1 depicts the random part of the four class model with random slopes (Model IIId) using standard latent class parameters. As can be seen, the four latent classes show different time patterns. The largest class one is most against abortion and class three is most in favour of abortion. Both latent classes are very stable over time. The overall level of latent class two is somewhat higher than that of class one, and it shows somewhat more change of attitude over time. People belonging to latent class four are very unstable: at the first two time points they are similar to class two, at the third time point to class four, and at the last time point again to class two. Class four could therefore be labelled as random responders. It is interesting to note that in a three class solution the random-responder class and class two are combined. Thus, by going from a three to a four class solution one identifies the interesting group with less stable attitudes.

Conclusions

In this paper we propose using the LC regression model as a tool for random coefficient modelling. We show how to transform the standard LC regression parameters into multilevel parameters, yielding the same type of insight into the random structure as with a parametric random coefficients model. The empirical example showed that the assumption of multivariate normality of the random coefficients may sometimes be too restrictive: the LC models fitted much better and detected the random slopes.

An important advantage of the non-parametric approach not yet mentioned is the much shorter computation time. Actually, the abortion example is a small problem for the Latent GOLD program: estimation of the largest model (IIIe) took only 3 seconds. In contrast, the estimation of the parametric model with 4 random coefficients took 18 minutes.

6. References

Agresti A., Booth, J. G., Hobert, J. P., and Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology*, **30**: 27-80.

Bryk, A. S., and Raudenbusch, S. W. (1992). *Hierarchical linear models: application and data analysis*, Newbury Park, CA: Sage.

Goldstein, H. (1995). *Multilevel statistical models (2nd ed)*, London: Arnold.

Hedeker, D., and Gibbons. R. D. (1996). MIXOR: A computer program for mixed effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine*, **49**: 157-176.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association*, **73**: 805-811.

McGrath, K., and Waterton, J. (1986). *British social attitudes, 1983-1986 panel survey*. London: Social and Community Planning Research, Technical Report.

Rabe-Hesketh, S., Pickles, A., and Skrondal, A. (2001). GLLAMM: A general class of multilevel models and a Stata program. *Multilevel Modelling Newsletter*, **13** (1): 17-23.

Vermunt, J. K., and Magidson, J. (2000). *Latent GOLD User's guide,* Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K. (1997). *LEM: A general program for the analysis of categorical data, Users' manual*, Tilburg University

Wedel, M., and DeSarbo, W. (1994). A review of recent developments in latent class regression models. In Bagozzi, R. P. (Eds.), *Advanced methods of marketing research*, 352-388. Cambridge, Massachusetts: Blackwell Publishers.

Using Mplus ® (Version 2.01) for multilevel modelling

Ian Plewis

Institute of Education, University of London

This is not a full review of the Mplus package which, as the manual's sub-title suggests, is mainly concerned with statistical modelling using latent variables. Rather, I focus on its utility as an additional tool for the multilevel modeller's toolbox that already contains one of the specialist packages.

Mplus has been developed and written by a team headed by Linda and Bengt Muthén. It comes as a set of three disks together with an extensive, well-produced manual. It is essentially a DOS package, based around a core of nine commands, each with associated sub-commands. There is no 'help' facility within the package but some exemplar analyses can be found on the company web site – http://www.statmodel.com. The manual also contains an extensive set of templates for particular models although these are a little unsatisfactory as they are all artificial with no substantive motivation, have no data attached to them, and hence they do not give any indication of what results might be expected and what they would look like.

Turning to the package's multilevel modelling capabilities, we find a 'twolevel' procedure. This is restricted not only to two levels but also just to random intercepts models and I did not find natural its way of setting up the model in terms of 'within' and 'between' components. The example on the web site is not especially helpful in this regard as its output appears to be restricted to just the 'between' results.

Probably the most useful part of the package for a multilevel modeller is the one that deals with growth curve modelling of longitudinal data. Mplus uses an approach where the repeated measures on a univariate outcome are seen as a multivariate outcome vector. So longitudinal data that are treated as two level in, for example, *MLwiN* are thought of as regular (single level) data in Mplus. Here, I was interested to see whether the package was able to solve some problems I had when using a structural equations approach to growth curves, treating the polynomial components of the curve as latent variables. Essentially I had been fitting a quadratic to two related growth processes, with up to six measurements for each process, some time-invariant covariates and the subjects nested within schools. (These are the data used in Plewis (2001).) I had wanted but had not previously been able to:

a) include 'school' as a third level, above occasions and pupils, both to allow for variability at that level and also to see whether the relations between the two processes varied across schools. In principle, by combining the two level approach with the growth curve approach, Mplus should at least make it possible to purge the between pupil covariance matrix of its between school components. (Remember that 'twolevel' does not permit random slopes.) However, the package only permits using this combination with listwise deletion for missing data, a serious drawback for most longitudinal data that suffer from attrition over time. Consequently, I was not able to make any progress on this front.

b) see whether the relations between the two growth processes varied with the values of the time-invariant level-two covariates. The seemingly obvious solution to this problem within Mplus – using multiple group analysis – allowed me to estimate different models for the sub-groups defined by my categorical level-two covariate but it did not appear to be possible to test the equality of the structural coefficients across these sub-groups. However, it is possible to tackle this problem via a mixture model, treating the covariate as a latent class variable that has a one-to-one correspondence with the categories of the background variable. I was then able to test for the equality of the structural coefficients, allowing for between-group differences in the estimated means and variances of the observed measures across time. (Incidentally, the results were not entirely conclusive, depending on what measure of fit – deviance, AIC, BIC - was chosen.) So I did find the package useful for this problem.

There are other possibilities for modelling longitudinal data within the package – for example, fitting growth curves to data from multiple cohorts, growth models for binary and ordered categorical outcomes, and discrete time survival analysis. It is also possible to analyse continuous responses produced by complex sample designs by incorporating weights into the analysis but the approach adopted is kept separate (both in the package and in the manual) from the multilevel modelling capabilities.

There were aspects of the estimation methods that caused me some concern. For example, no warning was given about negative estimated variances nor about inadmissible covariance matrices with correlations far in excess of one. It would also be useful to be able to increase from three the number of decimal places in the output.

To sum up, Mplus offers those multilevel modellers with longitudinal data some additional features that they might not find elsewhere. However, potential users might want to take into account some of the criticisms and concerns noted above.

Reference

Plewis, I. (2001). Explanatory models for relating growth processes. *Multivariate Behavioral Research*, **36**, 207-225.

Comment on Mplus review

Linda and Bengt Muthén
Muthén & Muthén

We would like to comment on the Plewis review of the multilevel modelling capabilities of Mplus, which as the reviewer points out is only a small part of the Mplus modelling capabilities. We believe that the review misses important longitudinal modelling features and misunderstands some other issues.

One point that is not clear from the review is that Mplus can estimate two level longitudinal models with missing data. This is done using maximum likelihood estimation with standard MAR assumptions. Random slopes are allowed. In addition clustering and weights can be taken into account. Clustering is taken into account by correcting standard errors using the sandwich estimator.

The reviewer was, however, correct that Mplus currently cannot include missing data in a three level analysis. This facility exists in a developmental version of Mplus for later release. Note, however, that the current version can be used to give corrected standard errors of parameter estimates taking into account both clustering and missing data for data that has three levels by using TYPE = MIXTURE MISSING COMPLEX in Mplus terms, where corrected standard errors can be obtained as described above.

An important omission of the review is that it does not mention the unique growth mixture modelling capabilities that Mplus offers for two level longitudinal modelling, including missing data and clustering. In growth mixture modelling, the conventional approach of using random coefficients for growth modelling is greatly enhanced by the ability also to allow for latent trajectory classes. The information that can be gained by using growth mixture modelling with latent trajectory classes is exemplified in a series of recent publications. See e.g. Muthén (2001a, b), Muthén and Shedden (1999), and Muthén et al. (in press). Again, corrected standard errors under cluster sampling can be obtained.

There are also a couple of misunderstandings.

(1) The reviewer states that he did not find full output for multilevel analysis on the Mplus web site. He was, however, not looking at the examples provided for multilevel modelling but an advertising section.

(2) He was also unable to use multiple group multilevel modelling to test equalities across groups. Such testing is, however, possible and straightforward.

(3) Weights are available for multilevel analysis.

References

Muthén, B. (2001a). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In Collins, L. M. and Sayer, A. (Eds.), *New Methods for the Analysis of Change*, 291-322. Washington, DC: APA.

Muthén, B. (2001b). Latent variable mixture modeling. In Marcoulides, G. A., and Schumacker, R. E. (Eds.), *New Developments and Techniques in Structural Equation Modeling* , 1-33. Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, B., and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, **55**: 463-469.

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., Wang, C. P., Kellam, S., Carlin, J., and Liao, J. (in press). General growth mixture modeling for randomized preventive interventions. *Biostatistics (forthcoming)*.

Review of 'Multilevel Modelling of Health Statistics'. Alastair Leyland & Harvey Goldstein (Eds.). Pp xvii & 217 (2001). New York: Wiley.

Bob Prosser

Pharmaceutical Outcomes Program, Children's & Women's Health Centre of British Columbia, Canada.

During recent years, growing numbers of health care researchers have discovered and contributed to what has become an impressive body of theoretical and practical knowledge concerning multilevel modelling. Although some multilevel applications within health care parallel the early work in the field of education that popularised the basic techniques—tracking performance of practitioners (students) within hospitals (schools), for example—the rich variety of forms of hierarchically structured data arising from health-related studies has led to a considerable broadening of the methodology's scope. (As Nigel Rice notes in chapter three, most data sets from health research "do not lend themselves to simple model specification allowing a linear link function to relate a set of explanatory variables to a response measured on a continuous scale.") The publication of *Multilevel Modelling of Health Statistics* fittingly marks a decade of rapid progress by showcasing many advanced models and statistical tools in a very interesting and (for the most part) highly readable manner.

The book contains 13 chapters most of which feature at least one case study to illustrate the application of specialised models or techniques within a wide spectrum of health studies. The concluding chapter is a review by Jan de Leeuw and Ita Kreft of the major computer programs used in performing these analyses.

The first chapter, by Michael Healy, provides a good but brief introduction to basic variance components and random slopes models using data about children's coughing and sleeping over several nights. Healy, like many of the authors, offers helpful explanations of statistical judgments made in the course of selecting and elaborating a model, and interpreting results. He mentions, for example, why he prefers to examine loglikelihoods in choosing between models instead of relying on the comparison of a random slope's variance with its standard error. Such pointers, woven in throughout the book, will be very useful to novice and experienced multilevel data analysts alike.

Harvey Goldstein and Geoff Woodhouse discuss models that are useful in the analysis of data from longitudinal studies. Their case study begins with a discussion of basic two level growth curves for adolescent boys' heights and an extension into a multivariate repeated measures model. The authors explain how to use this model for prediction. Key sections describe and illustrate the modelling of autocorrelated level-one random terms. Researchers who conduct clinical trials will appreciate the discussion of studies with crossover designs.

Rice's discussion of modelling binary response variables and proportions provides one of the book's more detailed expositions of theory. He includes an informative section on extra-binomial variation and model diagnostics and a very thorough description of all aspects of his interesting case study on equity in utilization of health care.

The chapter by Ian Langford and Rosemary Day begins with a discussion of when to use Poisson regression to model count data then succinctly outlines the associated theory. Their first example involves regional and international variation in testis cancer mortality (in relation to income and population density), and their second example examines weekly numbers of reported cases of food poisoning in England and Wales. In both examples the authors estimate higher level residuals to determine whether there are unusual regions.

Alice McLeod's chapter provides a good follow-up to the contributions by Healy and Rice, illustrating the extension of univariate multilevel modelling to multilevel multivariate regression—first for continuous response variables and second for discrete outcomes. This reviewer was particularly interested in the second example analysis in which hospital length of stay and emergency readmissions were modelled together (i.e., a continuous and a binary outcome) as functions of several covariates. In this case, the positive hospital level correlation between the two outcomes contradicted the hypothesis that shorter hospital stays may increase the risk of readmission. McLeod also provides one of the book's several discussions of treating missing data.

Disentangling the effects of different parts of a health care system on patient outcomes is a complex undertaking. A particular patient may be treated by more than one doctor for the same health problem, for example. Two patients receiving tertiary care in a hospital may well have received primary care in different health regions. Jon Rasbash and William Browne introduce readers to a systematic approach to conceptualising and modelling datasets involving such non-hierarchical health care system structures - in particular, cross-classified and multiple membership structures. Their interesting dual crossed hierarchy artificial insemination example whets the appetite for the additional details supplied in the referenced papers.

Many outcome variables in health care research are categorical (nominal or ordinal) with several categories, e.g., a discharge disposition variable indicating the type of care planned for a patient after an intervention. Min Yang's chapter discusses the multilevel modelling of such responses using the multinomial distribution for the level-one random terms. There are two three level case studies—both very thoroughly presented—the first involving a nominal variable for physician antibiotic prescribing practice and the second involving an ordinal tobacco and health knowledge score.

Spatial modelling extends the usual multilevel analysis of events within geographic regions by accounting for inter-regional proximity. Leyland's chapter on spatial analysis focuses on disease and mortality mapping using an autoregressive error structure. Effects of improving models in a case study of lip cancer incidence in Scottish districts are well illustrated via a series of shaded maps. Brief descriptions of multivariate spatial models and spatio-temporal models round out the discussion.

Two of the chapters provide technical advice and illustrations with particularly broad applicability. Toby Lewis and Ian Langford discuss identification and treatment of outliers, points of high leverage and influential points in multilevel analysis, and Tom Snijders addresses multilevel sampling issues such as level of randomisation, choice of covariates and determination of sample sizes at different levels. Both contributions are very helpful.

Issues concerning the comparison and explanation of performance of institutions or individual practitioners or both have been widely discussed from a multilevel modelling perspective within the domain of education. Clare Marshall and David Spiegelhalter examine performance of surgeons within hospitals in terms of risk-adjusted mortality rates, re-illustrating the importance of quantifying (lack of) precision in ranks and stressing caution in league table interpretation. Among the unique features of this chapter are the use of a Bayesian approach and the inclusion of the script used in analysis with the BUGS software.

Goldstein and Leyland round out the modelling presentations with a 'Further Topics' chapter discussing multilevel meta analysis, survival analysis and contextual analysis. While the authors convey key information clearly, each of these topics is worthy of a chapter and a full case study.

Although the Introduction describes the book as a "self-contained reference for graduate and higher level course for those with a knowledge of basic regression modelling," this reviewer imagines that many readers with just such a preparation will be consulting other books and papers to obtain further background. Fortunately, the reference list is extensive. The book begs for a companion workbook with data on a CD ROM to guide readers in using the major software packages to perform the illustrated (and other) analyses. All in all, however, this is a very valuable reference for experienced health care statisticians interested in designing studies and analysing results within a multilevel framework.

Some Recent Publications Using Multilevel Models

Albert, P. A., and Follman, D. A. (2000). Modeling repeated count data subject to informative dropout. *Biometrics*, **56**: 667-677.

Ansari, A., and Jedidi, K. (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika*, **65**: 475-496.

Belfield, C. R., and Fielding, A. (2001). Measuring the relationship between resources and outcomes in higher education in the UK. *Economics of Education Review*, **20**: 589-602.

Browne, W. J., Goldstein, H., and Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling* **1**: 103-124.

Carr-Hill, R. (2000). Developing a robust resource allocation formula for police. *Policing and Society*, **10**: 235-261.

Dominici, F., Samet, J. M., and Zeger, S. L. (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy (with discussion). *Journal of the Royal Statistical Society*, *A*, **163**: 262-302.

Fielding, A. (2001). Scaling for residual variance components of ordered category responses in generalised linear mixed multilevel models. In Klein, B. and Korsholm, L, (Eds.), *New Trends in Statistical Modelling*, Procs of 16th International Workshop on Statistical Modelling, 191-200.

Fitzmaurice, G. M., Laird, N. M., and Shneyer, L. (2001). An alternative parameterisation of the general linear mixture model for longitudinal data with non-ignorable dropouts. *Statistics in Medicine*, **20**: 1009-1022.

Fox, J., and Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, **66**: 271-288.

Friedman, L. (2000). Estimators of random effects variance components in meta analysis. *Journal of Educational and Behavioral Statistics*, **25**: 1-12.

Gill, P. S. (2000). A robust mixed linear model analysis for longitudinal data. *Statistics in Medicine*, **19**: 975-987.

Gray, J., Goldstein, H., and Thomas, S. (2001). Predicting the future: the role of past performance in determining trends in institutional effectiveness at A level. *British Educational Research Journal*, **27**: 391-406.

Greenland, S. (2000). When should epidemiologic regressions use random coefficients? *Biometrics*, **56**: 915-921.

Gueorguieva, R. V., and Agresti, A. (2001). A correlated probit model for joint modelling of clustered binary and continuous responses. *Journal of the American Statistical Association*, **96**: 1102-1112.

Hoshino, S. (2001). Multilevel modeling on farmland distribution in Japan. *Land Use Policy*, **18**: 75-90.

Hox, J., and Maas, C. J. M. (2001). The accuracy of multilevel structural equation modelling with pseudobalanced groups and small samples. *Structural Equation Modelling*, **8**: 157-174.

Ibrahim, J. G., Chen, M. H., and Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is non-ignorable. *Biometrika*, **88**: 551-564.

Jung, S. H., Kang, S. H., and Ahn, C. (2001). Sample size calculations for clustered binary data. *Statistics in Medicine*, **20**: 1971-1982.

Lenk, P. J., and DeSarbo, W. S. (2000). Bayesian inference for finite mixtures of generalised linear models with random effects. *Psychometrika*, **65**: 93-119.

Littell, R. C., Pendergast, J., and Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine*, **19**: 1793-1820.

Longford, N. (2000). On estimating standard errors in multilevel analysis. *Journal of the Royal Statistical Society*, *D*, **49**: 389-398.

Malin, A. and Linnaklya, P. (2001). Multilevel modelling in repeated measures of the quality of Finnish school life. *Scandinavian Journal of Educational Research*, **45**: 145-166.

Merlo, J., Östergren, P.-O., Lindström, M., Lindgren, A., Melander, A., Råstam, L., Berglund, G. (2001). Diastolic blood pressure and area of residence: multilevel versus ecological analysis of social inequality. *Journal of Epidemiology and Community Health*, **55**: 791-798.

Moerbeek, M. (2000). *Design and analysis of multilevel intervention studies*. Maastricht: Maastricht University.

Moerbeek, M., Van Breukelen, G. J. P., and Berger, M. P. F. (2000). Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics*, **25**: 271-284.

Moerbeek, M., Van Breukelen, J. P. and Berger, M. P. (2001). Optimal experimental design for multilevel statistical models. *Journal of the Royal Statistical Society, D,* **50**: 17-30.

Nelder, J. and Lee, Y. (2000). *Hierarchical GLMs*. London: Imperial College.

Omar, R. Z. and Thompson, S. G. (2000). Analysis of a cluster randomised trial with binary outcome data using a multilevel model. *Statistics in Medicine*, **19**: 2675-2688.

Polsky, C., and Easterling, W. E. (2001). Adaptation to climate variability and change in the US Great Plains: A multi-scale analysis of Ricardian climate sensitivities. *Agriculture, Ecosystems, and Environment*, **85** (1-3): 133-144.

Rodriguez, G., and Goldman, N. (2001). Improved estimation procedures for multilevel models with binary response: a case study. *Journal of the Royal Statistical Society*, *A*, **164**: 339-356.

Rowe, K. J. (2001). *Multilevel structural equation modelling with MLn/MLwiN & Lisrel 8.30*. Melbourne: Autralian Council for Educational Research.

Special Issue on Multilevel Models (2001). *Multivariate Behavioral Research*, **36** (2).

Spiegelhalter, D. (2001). Bayesian methods for cluster randomised trials with continuous responses. *Statistics in Medicine*, **20**: 435-452.

Ten Have, T. R., Miller, M. E., Reboussin, B. A., and James, M. K. (2000). Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple-cause drop-out from the longitudinal study of aging. *Biometrics*, **56**: 279-287.

Turner, R. M., Omar, R. Z., Yang. M., Goldstein, H. and Thompson, S. G. (2000). Random effects meta-analysis with binary outcomes using multilevel models. *Statistics in Medicine*, 19: 3417-3432.

Turner, R. M., Omar, R. Z., and Thompson, S. G. (2001). Bayesian methods of analysis for cluster randomised trials with binary outcome data. *Statistics in Medicine*, **20**: 453-472.

Vaida, F., and Xu, R. (2001). Proportional hazards model with random effects. *Statistics in Medicine* **20**: 3309-3324.

Visscher, A. (2001). Public school performance indicators: problems and recommendations. *Studies in Educational Evaluation*, **27**: 199-214.

Willms, J. D. (2000). *Three hypotheses about community effects relevant to the contribution of human and social capital to sustaining economic growth and well-being.* University of New Brunswick.

Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society, C*, **50**: 375-388.

Yang, M., and Woodhouse, G. (2001). Progress from GCSE to A and AS level: institutional and gender differences, and trends over time. *British Journal of Educational Research*, **27** (3): 245-267.

Yang, M., Goldstein, H., Browne, W. J. and Woodhouse, G. (2002). Multivariate multilevel analysis of examination results. *Journal of the Royal Statistical Society, A*, **165** (1).

Yau, K. K. W. (2001). Multilevel models for survival analysis with random effects. *Biometrics*, **57**: 96-102.

Zhang, D., and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, **57**: 795-802.

Please send us your new publications in multilevel modelling for inclusion in this section in future issues.

MLwiN User Survey

We have conducted a survey of our users to find out what *MLwiN* is being used for and how we can improve the package. Thanks to all of you who responded to the survey. We are in the middle of analysing the comments and formulating our responses. We will publish the results of this survey, along with our reaction, in the next issue of the newsletter.

For those who have not yet participated in the survey, we would still welcome your views. The form is at: http://multilevel.ioe.ac.uk/support/surveyfm.html